FORECASTING OF EVAPORATION FROM HEMREN RESERVOIR BY USING ARTIFICIAL NEURAL NETWORK

SAAD SH. SAMMEN

Assistant Lecturer, College of Engineering, Diyala University, Iraq.
E-mail: saad123engineer@yahoo.com
(Received: 12/9/2012; Accepted:4 /11/2012)

ABSTRACT: - The evaporation is one of the basic components of the hydrologic cycle and is essential for studies such as water balance, irrigation system design, and water resource management and it is requires knowledge of the values of many climatic variables. In order to estimate the evaporation, direct measurement methods or physical and empirical models can be used. Using direct methods require installing meteorological stations and instruments for measuring evaporation. Installing such instruments in various areas requires specific facilities and cost which is hard to be employed. Accordingly, this paper is an attempt to assess the potential and usefulness of ANN based modeling for evaporation prediction from Hemren reservoir by using daily temperature, relative humidity, wind velocity, sunshine hours, and evaporation data in Hemren meteorological station. Also, this study outlines a procedure to evaluate the effects of input variables on the output variable using the weight connections of ANN models.

The Lev. Marqn. Back Prog. (LMBP) has been utilized to construct the ANN models. For the development ANN model, different networks with different numbers of neurons and layers were evaluated. Mean Squared Error (MSE) and the Correlation Coefficient (R^2) were employed to evaluate the accuracy of the proposed model. The study shows that the best model for estimation of evaporation is ANN (4-10-1), it have MSE equal to 0.112711 and the correlation coefficient (R^2) equal to 0.999540.

Keywords: Artificial neural networks, Daily Evaporation, Hemren Reservoir.

1- INTRODUCTION

The decrease of water availability over the last few decades is one of the principal environmental problems that could severely restrict agricultural and industrial development

of some arid and semiarid areas of the world. In agriculture, water requirements are associated with irrigation use. Prediction of irrigation water demand involves a computation of many water balance factors such as runoff, ground water table contribution, precipitation and evaporation.

Evaporation refers to water losses from the surface of a water body to the atmosphere and occurs when the number of moving molecules that break from the water surface and escape into the air as vapor is larger than the number that re-enters the water surface from the air and become entrapped in the liquid. Evaporation reflects the influence of several meteorological parameters like air temperature, sunshine hours, wind velocity, relative humidity, solar radiation, evaporating power of the air and vapor pressure deficit of a locality, where evaporation increases with high wind speed, high temperatures and low humidity. All of this factors caused that be evaporation estimation one of the most complexity calculations in meteorology and hydrology sciences. Measurement of evaporation is very complicated. In meteorological stations used class pan evaporation, Measurement of evaporation by this pan has limitations and it is an index for evaporation of region.

A sizable quantity of water is lost every year by evaporation from storage reservoirs and evaporation of water from large water bodies influences the hydrological cycle. Among the hydrological cycle, evaporation is perhaps the most difficult to estimate due to complex interactions among the components of land-plant-atmosphere system (Singh 1997).

The literature review showed that these methodologies vary greatly in their ability to define the magnitude and variability of the evaporation from the reservoirs. It is therefore necessary to develop alternate approaches to estimate the evaporation rates based on metrology variables, which are comparatively easier to measurements and estimation. One of the recent alternate approaches is the use of soft computing modeling techniques, which have better modeling flexibility and capability rather than previous empirical approaches where in this modeling each of the metrological parameter takes it share proportionally.

2. MATERIALS AND METHODS

2-1. Description of the Study Area

The Hemren Lake lies in the northeastern part of Iraq, about 120 Km from the Baghdad city within Diyala governorate (Figure1). The lake has a full–pool operating altitude of 92 m above mean sea level and its boundaries extend between a latitude of 25º61' 34º14' N and a longitude of 30º12'- 44º09'E. It was formed during 1976-1981 by Hemren dam on the

Diyala River. Morphometric features of Hemren lake indicate that: the volume of the lake is 2.06 billion cubic meters with a surface area of about 327 km^2 at high level period. The area faced two successive drought years (2008 and 2009).

2-2. Description of data used

The average weekly evaporation and meteorological data of Hemren reservoir (from 2008 to 2012) were collected by (Hemren dam project management) as mentioned in (Table1). The measured meteorological variables include hourly observations of evaporation, mean air temperature; mean sunshine hours, mean relative humidity, and mean wind velocity that were used in the present study.

3 - ARTIFICIAL NEURAL NETWORKS MODELING

3-1. Artificial Neural Network

ANN was first introduced as a mathematical aid by McCulloch and Pitts (Mcculloch 1943). They were inspired by the neural structure of the brain. (Figure 2) shows the structure of a commonly used back propagation neural network. A neural network is an artificial intelligence technique that mimics a function of the human brain. It is a capable of identifying complex nonlinear relationships between input and output data sets. ANN models have been found useful and efficient, particularly in problems for which the characteristics of the processes are difficult to describe using physical equations. The conventional model requires many input parameters and variables, some of which cannot be obtained easily from the field and may vary from site to site even within the same geographical region. An artificial neural network has found successful applications in the areas of science, engineering, industry, business, economics and agriculture. For example, artificial neural networks have improved the technique of satellite images and patterns recognition, wastewater treatment, remote sensing and seawater pollution classification. Neural Networks (NN) have been used to model a variety of biological and environmental processes by Altendorf (Altendorf and Stone, 1999) and (Gallant, 1993).

Most of the available models for estimation of evaporation may be reliable and most appropriate for use in climates similar to where they were developed. It is likely that errors may occur when these models are used under climatic conditions that are different under which they were developed.

The pan evaporation approaches reviewed generally performed better when solar radiation and relative humidity were included as input variables. However these data are often not available. Models of pan evaporation are required to be established that corresponds to the relatively minimum weather data variable for most of the locations. In earlier research work of empirical modeling and equation were fit to the data and the correlation was determined with same data set. No attempt was made to apply those models to the other independent data set of different location. The evaporation prediction also highly depends upon the quality of the pan evaporation measurements used in the ANN modeling.

3-2. ANN Structure

Most ANNs have three layers or more: an input layer, which is used to present data to the network; an output layer, which is used to produce an appropriate response to the given input; and one or more intermediate layers, which are used to act as a collection of feature detectors.

There are a wide variety of algorithms available for training a network and adjusting its weights. In this study, an adaptive technique momentum Levenberg–Marquardt based on the generalized delta rule was adopted.

Let x_i ($i = 1,2, ... n$) are inputs and w_i ($i = 1,2, … n$) are respective weights. The net input to the node can be expressed as

$$net = \sum_{i=1}^{n} x_i w_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(1)$$

The net input is then passed through an activation function f ( ) and the output (y) of the node is computed as

$$y = f(net)$$

Sigmoid function is the common transfer function which is the expressed by

$$y = f(net) = \frac{1}{1+e^{-net}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (2)$$

It's known that the increasing of the hidden layers number affects the complexity of the network and decreases the learning accuracy. Theoretical works have shown that a single hidden layer is sufficient for the ANNs to approximate to any complex nonlinear function (Cybenko, 1989; Hornik et al., 1989).

Meanwhile in 1991, Tang et al. recommended using one hidden layer to avoid increasing the complexity of the network. Indeed, many experimental results seem to confirm that one hidden layer may be enough for most forecasting problems (Coulibaly et al., 1999).

Each layer includes several neurons and the layers are interconnected by sets of correlation weights. The neurons produce outputs through transforming inputs comes from the initial inputs or the interconnections, by an adequate nonlinear transfer function.

Commonly, Sigmoid function used in back propagation networks (BPN). The training processing of neural network is essentially executed through a series of patterns, and for the learning process, the interconnection weights are adjusted within input and output values. The back propagation is the most representative learning model for the artificial neural network.

In the back propagation procedure, the error at the output layer propagates backward to the input layer via the hidden layer in the network obtaining the final optimum outputs.

3-3. Levenberg-Marquardt Back Propagation Algorithm

The ANN-based modeling was reviewed in the hydrology over the last years by Coulibaly et al. (1999) and reported that about 90% of the experiments, extensively make use of the multi-layer feed-forward neural networks (FNN) trained by the back propagation algorithm. Early stop training approach Levenberg-Marquardt back propagate on algorithm (LMBP) has widely been used for the hydrological forecasting problems, as the most powerful and accurate algorithm (Cobaner et al., 2008; Vongkunghae and Chumthong, 2007).

The Levenberg- Marquardt algorithm represents a simplified model of Newton's method applied to the problem of training multilayer perceptron neural networks.

3-4. Selection of ANN model input variables

Depending on the data of table (1), the correlation matrix between all the input variables as proposed by (Box and Jenkins, 1976) was presented in (Table 2). It can be observed from this table that the linear correlation between evaporation and temperature is 0.52, implying that any model built using radiation will certainly be able to compute the Evaporation satisfactorily (Jain et. 2008).

The model's accuracy can be improved by incorporating variables that account for aerodynamic effects on evaporation, such as humidity in addition to temperature. So, a combination of temperature and humidity may also provide a good estimate. However, it is to be noted that these inferences are based on the linear correlation between the variables, while the evaporation process is considered to be highly nonlinear. Hence, these inferences can be considered only as guidelines for deciding on the input variable combination. Consequently the current study investigated the five combinations listed in (Table 3).

3-5. ANN Model Construction and Evaluation

A standard back propagation algorithm was employed to estimate the network parameters (weights and biases). The activation function used in the hidden and output node is a sigmoid function. Determination of an appropriate architecture for a neural network (number of hidden nodes) for a particular problem is an important issue as the network topology directly affects its computational complexity and its generalization capability. Many studies have revealed that larger than-necessary networks tend to over-fit the training samples and thus have poor generalization performance, while too-small networks (that is, with very few hidden neurons) will have difficulty in learning the training data. Currently there is no established methodology for selecting the appropriate network architecture before training (Coulibaly et al., 2001).

In the present study, the number of hidden neurons in the network was identified by various trials. The trial and error procedure started with two hidden neurons initially, and the number of hidden neurons was increased to 10 with a step size of 2 at each trial. For each set of hidden neurons, the network was trained in batch mode to minimize the mean square error at the output layer. In order to check any over-fitting during training, cross-validation was performed by keeping track of the efficiency of the fitted model. Though the available data was divided into three sets (training, validation and testing). The training was stopped when there was no significant improvement in the efficiency.

The values of evaporation computed by all the models were analyzed statistically using various indices employed for the performance analysis of models. The goodness-of fit statistics considered are the coefficient of correlation (R^2) between ANN computed and targeted evaporation and the model efficiency (Nash and Sutcliffe, 1970). Based on this analysis, the best architecture for each ANN model was identified. These selected models were subjected to further evaluation for their effectiveness in estimating the evaporation.

4. RESULTS AND DISCUSSION

4-1. Model Architecture Selection

The results, in terms of the performance indices for the trial and error procedure employed to identify the appropriate network architecture for each input combination, are shown in (Table 4). It is evident from this table that as the number of hidden neurons increases, the model performance also increases, up to a certain level. Thereafter, any additional neuron in the hidden layer dampens model performance. It is observed that for

Model 1, the correlation and efficiency statistics during training increase along with the number of hidden neurons. However, the efficiency is found to reach to acceptable value for ten neurons in the hidden layer, suggesting that the model is trying to over-fit the data. Therefore, ten hidden neurons were considered to be appropriate for Model 1. In a similar way, the optimal number of hidden neurons was determined for the other models which found equal to ten. Finally, it founded that the model 5 is consider appropriate model for evaporation estimation from Hemren reservoir.

4-2. Performance Evaluation of Models

The first set of analysis was carried out by using ANN with input as temperature and output as evaporation loss and results are plotted as shown in (Figure 3). The ANN determined a relationship (build a model) between the input and the output of the available data set of any system. These models are then used to predict the output from the known input values of the same system, thus requiring sufficient number of data to build and test the models. Similarly (Temperature, Humidity), (Temperature, Humidity, Wind velocity) and (Temperature, Humidity, sunshine hours) were taken as input parameters separately and evaporation loss were predicted and the results obtained are plotted as shown in (Figures 4-6) respectively. (Figure 7) was obtained when all four parameters were taken for model building and evaporation loss was predicted. Diagrams between actual and predicted daily pan evaporation values are plotted considering different input combinations of meteorological Parameters. The results of the different models obtained by using ANN are compared with the actual pan evaporation values of the reservoir. The examination of plotted (Figures 3-7) shows that there is better agreement of predicted results based on ANN.

The maximum value of correlation coefficient was obtained for model 5 (r=0.999540) followed by model 4 (R^2 = 0.99921), model 3 (0.998374), model 2 (R^2 = 0.987213) while minimum value was obtained for model 1 (R^2 = 0.920122). The performance parameters namely correlation coefficient and MSE obtained by using ANN (Table 4-8) are plotted in (Figures 8-9), while considering Temperature (minimum correlation coefficient = 0.92 and maximum MSE = 22.72) and all parameters (maximum correlation coefficient = 0.99 and minimum MSE = 0.11).

5- CONCLUSION

ANN has been proposed and emerged as an alternative approach of evaporation estimation from a Hemren reservoir. The back propagation multilayer perception ANN based modeling techniques when different parameters are used as input for model building for the prediction of evaporation. Further, the critical examination of plotted figures indicates that the performance of the ANN modeling was depends to a great extent.   The study also concludes that combination of all input parameters provides better performance of model for estimation the evaporation rather than individual parameters. The outcome of the study provided an impetus to the potential use of ANN approach for prediction the evaporation from the reservoirs in water resources projects.

References

1- Altendorf, C.T., Elliott, R. C., Stevens, E. W. and Stone, V M. L., (1999), "Development and validation of a neural network model for soil water content prediction with comparison to regression techniques", Trans., ASAE, vol. 42, no.3, pp691-699,.

2- Box GEP, Jenkins G.M., (1976), "Time Series Analysis: Forecasting and Control", Holden Day Inc.: San Francisco, USA.

3- Cybenko G., (1989) "Approximation by superpositions of a sigmoidal function", Math. Ctrl. Signals Syst. (MCSS), 2: 303–314.

4- Coulibaly P, Anctil F. Bobée B. (1999). "Prévision hydrologique parréseaux de neurones artificiels: état de l'art", Can. J. Civ. Eng., 26(3): 293-304.

5- Coulibaly, P., Anctil, F., Aravena, R., Bobee, B., (2001), "Artificial neural network modeling of water table depth fluctuations", Water Resources Research 37: 885–896.

6- Cobaner, M., Information, C., Haktanir, T., Kisi, O., (2008), "Prediction of Hydropower Energy Using ANN for the Feasibility of Hydropower Plant Installation to an Existing Irrigation Dam", Water Resour. Manage, 22: 757-774.

7- Hornik, K., Stinchcombe, M., White, H., (1989), "Multilayer feed forward networks are universal approximators. Neural Netw", 2: 359-366.

8- Gallant, S.T., (1993), "Neural Network learning and Expert systems", Cambridge, Mass : MIT press.

9- Jain, S. K., Nayak, P.C., and Sudheer, K.P., (2008), "Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation", Hydological Processes, 22 : 2225 – 2234.

10- Legates, D.R. and McCabe, G.J., (1999), "Evaluating the use of goodness-of-fit measures in hydrologic and hydro-climatic model validation", Water Resources Research, 35 (1): 233-241.

11- Mcculloch, W. S., and Pitts, W., (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics 5: pp.115-133.

12- Ministry of Water Resources, General Directorate of Dams and Reservoirs, Hemren Dam Project Management.

13- Nash, J.E, and Sutcliffe, J.V., (1970), "River flow forecasting through conceptual models: 1. A discussion of principles", Journal of Hydrology 10: pp.282–290.

14- Schaap, M.G. and Leij, F.J., (1998), "Database related accuracy and uncertainty of pedotransfer functions", Soil Science, 163: 765-779.

15- Singh, V. P. and Xu, C. Y., (1997), "Evaluation and generalization of 13 mass transfer equations for determining free water evaporation," *Hydrological Processes,* vol. 11, pp. 311-323.

16- Vongkunghae, A., Chumthong, A., (2007), "The performance comparisons of backpropagation algorithm's family on a set of logical functions", Trans. Electrical Eng., Electron. Commun, 5: 114-118.

Table (1): Average Weekly Metrological Data of Hemren Station from 2008 to 2012 (Hemren dam project management)

Month	Week	Evaporation Rate (mm/day)	Mean air Temperature (°C)	Average Wind Velocity (m/sec)	Average Sunshine Hours (hrs/day)	Mean Relative Humidity
Jan.	1	0.47	12.36	2.68	6.57	74.83
	2	0.51	16.64	2.21	5.91	57.96
	3	1.71	16.48	2.44	6.58	49.74
	4	0.73	10.48	2.44	6.68	65.23
Feb.	1	0.94	11.14	5.21	6.49	60.10
	2	3.49	16.58	6.27	7.13	54.14
	3	3.75	12.15	7.99	6.77	46.84
	4	7.64	11.16	6.59	6.46	46.66
March	1	2.08	11.99	11.60	7.96	50.86
	2	2.9	19.90	12.42	8.23	39.00
	3	2.27	18.19	12.02	9.10	32.89
	4	2.81	17.77	12.69	9.21	32.99
April	1	4.57	22.22	35.11	8.65	37.56
	2	3.84	22.94	35.11	8.58	41.04
	3	3.83	22.16	18.54	8.19	40.76
	4	4.69	25.68	10.09	8.30	35.33
May	1	4.86	27.56	25.71	7.68	35.13
	2	3.65	30.25	35.13	7.81	23.11

	3	3.68	31.30	15.65	7.33	22.69
	4	4.48	31.75	13.13	8.57	19.27
June	1	38.57	34.89	4.50	8.44	16.15
	2	34.81	34.69	3.00	6.53	18.40
	3	50.34	35.43	5.47	7.25	15.92
	4	6.39	35.68	2.85	7.85	18.59
July	1	9.3	35.73	4.88	4.79	17.75
	2	41.26	37.50	4.88	7.44	14.87
	3	16.4	36.45	4.88	7.04	18.85
	4	34.49	37.63	4.88	9.60	17.68
Aug.	1	5.56	38.86	4.08	10.69	14.23
	2	5.47	38.15	3.03	11.35	14.18
	3	4.44	37.57	4.13	11.43	15.74
	4	44.57	36.76	3.32	11.30	16.50
Sep.	1	6.3	35.07	4.14	10.64	20.29
	2	5.22	33.93	2.91	10.32	21.79
	3	4.75	31.88	2.80	9.93	24.27
	4	3.49	31.45	3.35	10.25	26.68
Oct.	1	3.47	26.47	2.97	9.95	33.66
	2	3.5	28.71	2.68	9.37	35.71
	3	1.82	26.73	3.69	9.21	29.68
	4	1.47	24.39	2.74	6.91	39.15
Nov.	1	1.13	19.35	2.84	7.68	41.01
	2	0.8	19.11	3.00	8.97	34.82
	3	1.95	17.43	2.75	6.20	43.49
	4	0.62	14.21	2.09	7.90	44.74
Dec.	1	0.78	14.21	2.90	8.13	35.40
	2	1.97	12.25	3.32	6.48	52.99
	3	0.79	13.40	3.73	6.65	61.11
	4	0.78	13.82	4.59	7.67	50.06

Table (2): Correlation factor between input and output variables.

	Evaporation	Temperature	Relative Humidity	Wind Velocity	Sunshine Hours
Evaporation	1.00				
Temperature	0.52	1.00			
Relative Humidity	-0.51	-0.92	1.00		
Wind Velocity	-0.1	-0.01	-0.03	1.00	
Sunshine Hours	-0.05	0.51	-0.53	0.03	1.00

Table (3): Combinations of input variables considered in developing ANN models.

Model	Input vectors
Model 1	Temperature
Model 2	Temperature, Humidity
Model 3	Temperature, Humidity, Wind speed
Model 4	Temperature, Humidity, sunshine hours
Model 5	Temperature, Humidity, Wind speed, sunshine hours

Table (4): Variation in performance of different models with number of hidden neurons in the ANN architecture in terms of correlation and efficiency statistics.

	No. of Neurons	(R)	MES
Model 1 (T)	2	0.690613	67.95734
	4	0.706828	53.32928
	6	0.760783	55.49758
	8	0.91485	24.11382
	10	0.920122	22.72411
Model 2 (T,H)	2	0.664909	39.05438
	4	0.844204	30.97595
	6	0.91068	19.43627
	8	0.955245	12.31883
	10	0.987213	3.2666
Model 3 (T,H,W)	2	0.941327	25.18595
	4	0.985999	1.21512
	6	0.989065	3.80553
	8	0.996989	0.318948
	10	0.998374	0.520210
Model 4 (T,H,SH)	2	0.98274	6.49284
	4	0.99064	301,007
	6	0.99567	0.97524
	8	0.99897	0.32943
	10	0.99921	0.16829
Model 5 (T,H,W,SH)	2	0.963163	7.73152
	4	0.992701	1.52216
	6	0.995707	0.638095
	No. of Neurons	(R)	MES
	8	0.997955	0.489956
	10	0.999540	0.112711

Figure (1): Hemren Lake Basin.
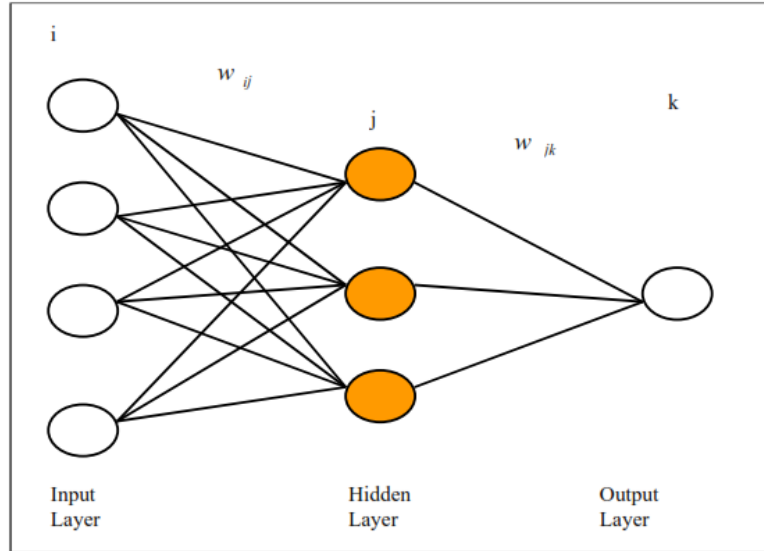
Figure (2): Structure of a back propagation neural network.
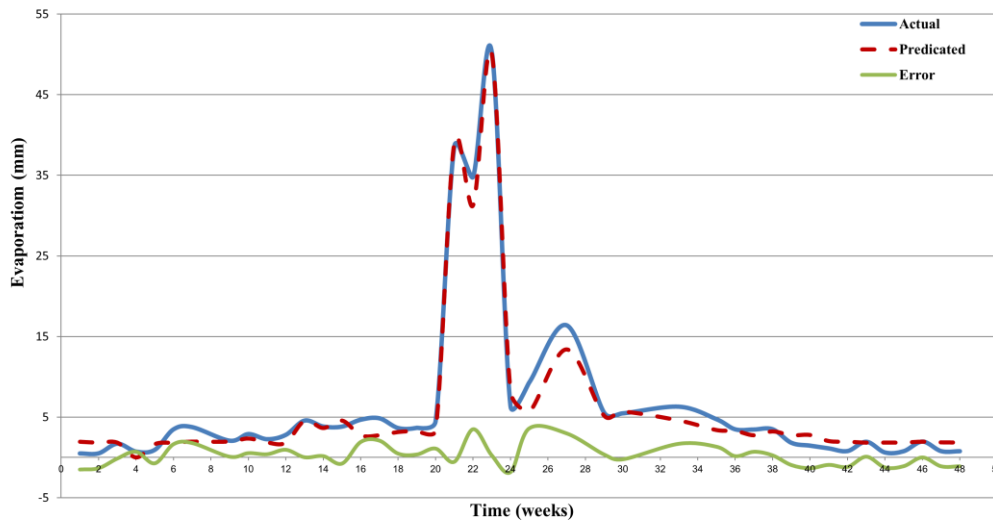
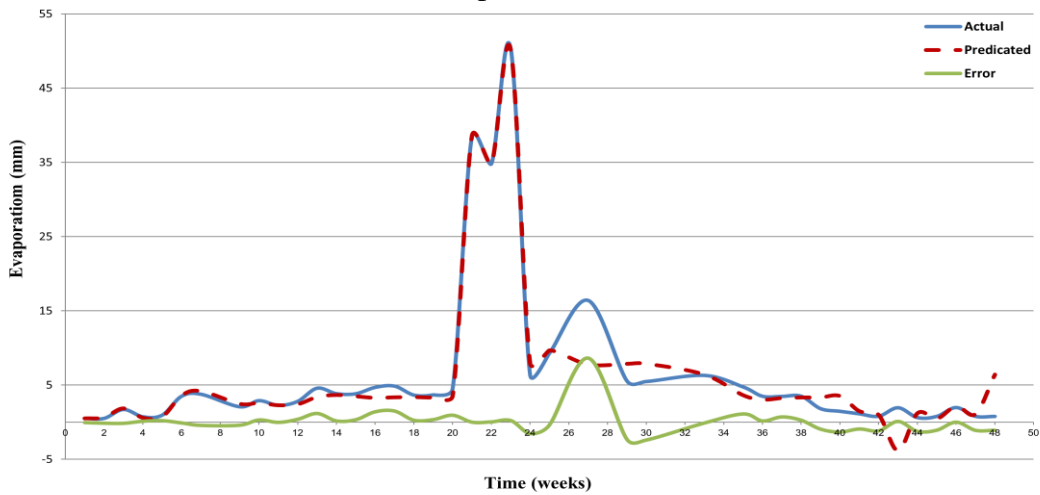**Figure (3):** Variation of actual evaporation versus predicted evaporation with temperature.

Figure (4): Variation of actual evaporation versus predicted evaporation with temperature and Humidity.
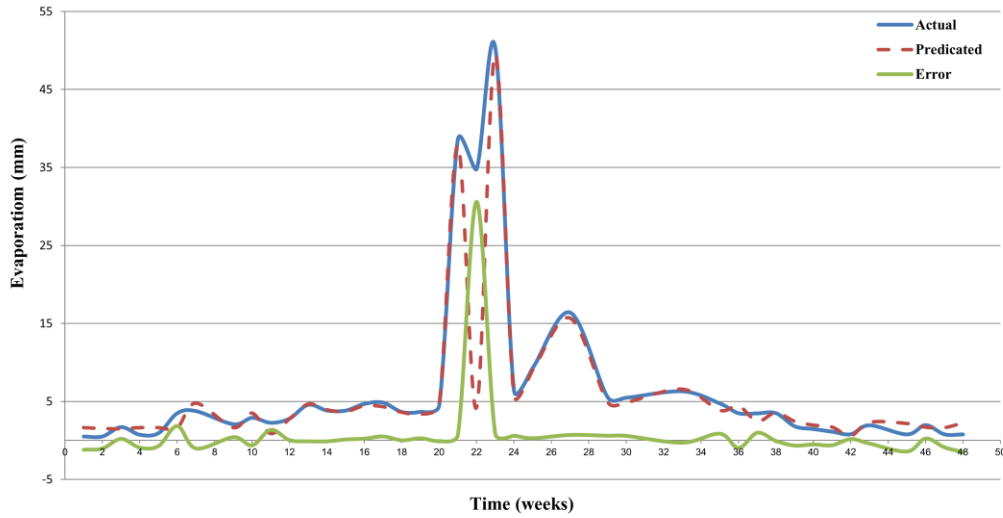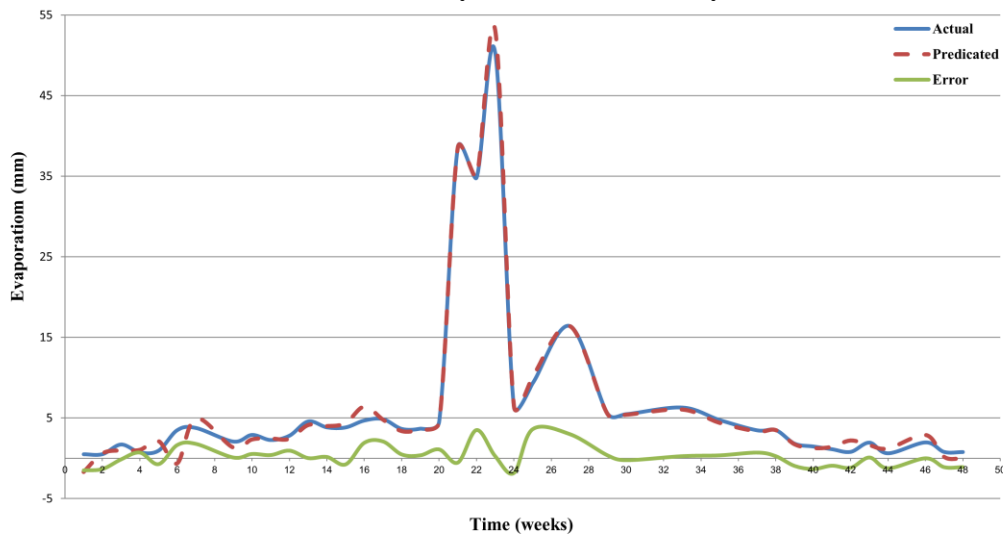
Figure (5): Variation of actual evaporation versus predicted evaporation with temperature, Humidity and Wind Velocity.

Figure (6): Variation of actual evaporation versus predicted evaporation with temperature, Humidity and Sunshine Hours.
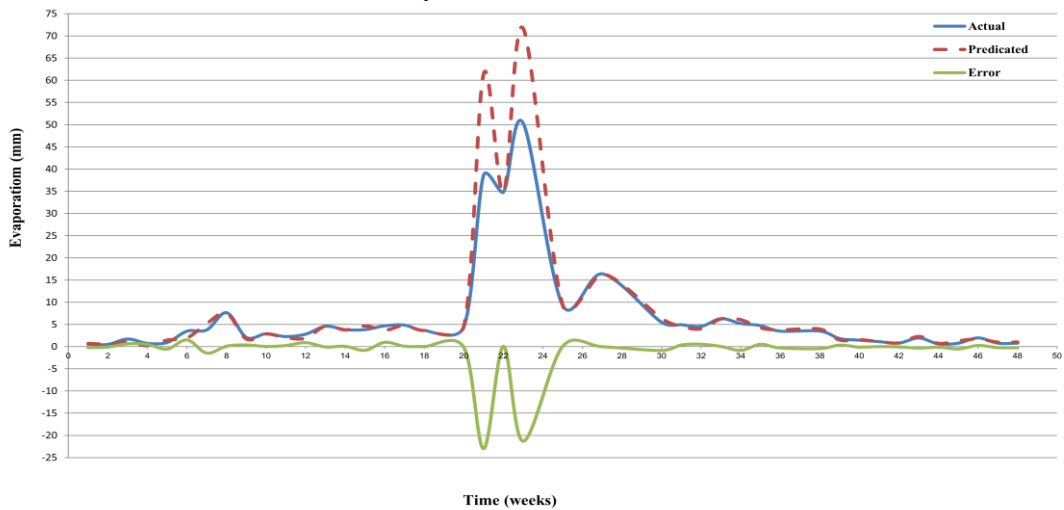

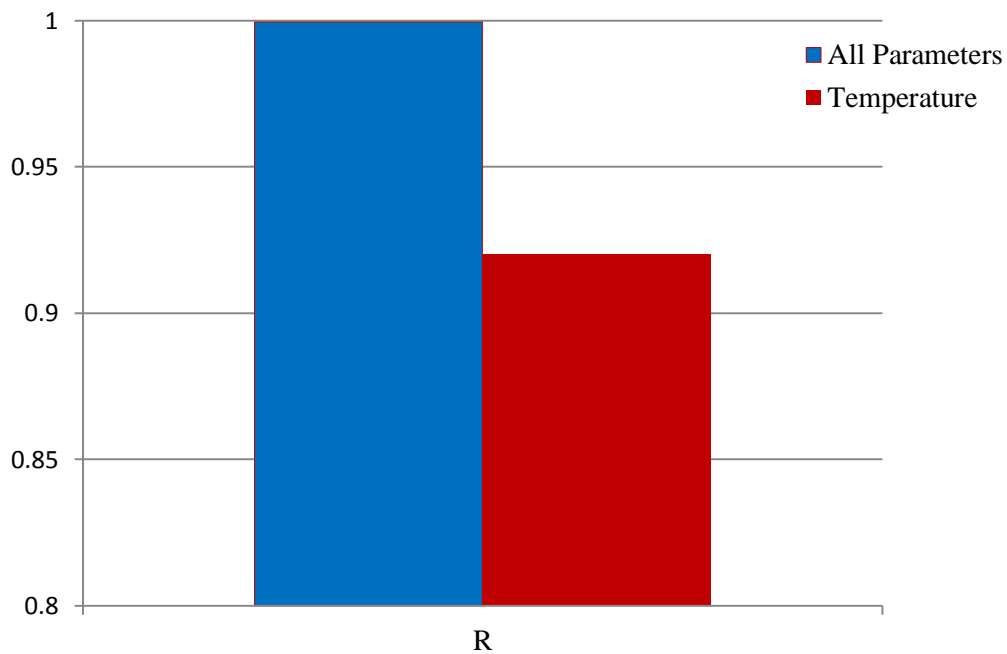**Figure (7):** Variation of actual evaporation versus predicted evaporation with all parameters.

Figure (8): Variation of correlation coefficient(R) between two models (Temperature and all parameters).
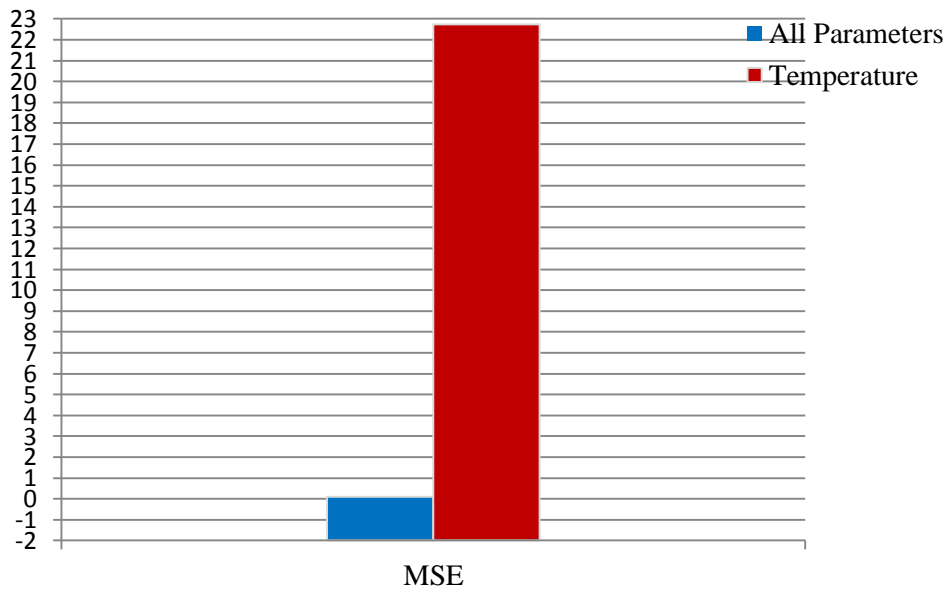

Figure (9): Variation of correlation coefficient (MSE) between two models (Temperature and all parameters).

التنبأ بالتبخر من خزان حمرين باستخدام الشبكات العصبية الصطناعية

سعد شوكت سمين

مدرس مساعد

كلية الهندسة/ جامعة ديالى/ العراق

الخلاصة

يعتبر التبخر احد العناصر الاساسية للدورة الهيدرولوجية وهو ضروري للحسابات التي تخص الموازنة المائية، تصاميم انظمة الري وادارة الموارد المائية وحسابه يتطلب معرفة العديد من المتغيرات المناخية.

هناك عدة طرق لحساب التبخر منها الرياضية والوضعية وطريقة الحساب المباشر. استخدام طريقة الحساب المباشر يتطلب نصب محطات الرصي الجوي ومعدات لقياس التبخر. نصب مثل هذه المعدات في مناطق مختلفة يتطلب كلفة ومتطلبات من صعب تحقيقها. لذلك هذا البحث هو محاولة لتقييم امكانية وكفاءة النماذج المعتمدة على الشبكات العصبية الاصطناعية لغرض توقع كمية التبخر في خزان حمرين باستخدام المعدلات اليومية للحرارة، الرطوية النسبية، سرعة الرياح، عدد ساعات النهار و كمية التبخر لمحطة حمرين الرصد الجوي.

لغرض بناء نموذج الشبكة العصبية الاصطناعية، عدة سبكات مختلفة وبعدد اعصاب مختلفة وطبقات مختلفة استعملت وتم اختيار الشبكة الافضل من بينها. تم اعتماد معامل الارتباط بين المعلومات الحقيقية والمعلومات التي الحصول عليها من النموذج الرياضي لغرض تقييم دقة النموذج المقترح، وكان مقدار معامل ارتباط لافضل نموذج (النموذج 1 – 10 – 4) هو (0.99954).